



Computing the Minimum-Phase Filter using the QL-Factorization

Hansen, Morten; Christensen, Lars P.B.; Winther, Ole

Published in:
IEEE Transactions on Signal Processing

Link to article, DOI:
[10.1109/TSP.2010.2045795](https://doi.org/10.1109/TSP.2010.2045795)

Publication date:
2010

Document Version
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

Citation (APA):
Hansen, M., Christensen, L. P. B., & Winther, O. (2010). Computing the Minimum-Phase Filter using the QL-Factorization. *IEEE Transactions on Signal Processing*, 58(6), 3195-3205.
<https://doi.org/10.1109/TSP.2010.2045795>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Computing the Minimum-Phase Filter using the QL-Factorization

Morten Hansen, Lars P. B. Christensen, and Ole Winther,

Abstract

We investigate the QL-factorization of a time-invariant convolutive filtering matrix and show that this factorization not only provides the finite length equivalent to the minimum-phase filter, but also gives the associated all-pass filter. The convergence properties are analyzed and we derive the exact convergence rate and an upper bound for a simple Single-Input Single-Output system with filter length $L = 2$. Finally, this upper bound is used to derive an approximation of the convergence rate for systems of arbitrary length. Implementation-wise, the method has the advantage of being numerically stable and straight forward to extend to the Multiple-Input Multiple-Output case. Furthermore, due to the existence of fast QL-factorization methods, it is possible to compute the filters efficiently.

Index Terms

Minimum-phase filtering, QL-factorization, sphere detection, spectral factorization, wireless communications.

I. INTRODUCTION

The minimum-phase and the all-pass filters have over the years attracted much attention due to their broad applicability in signal processing. For this reason these types of filters are generally covered in various classical books on signal processing, cf. e.g. [1]–[3]. One area where the minimum-phase filter is widely used is in digital communications over multipath channels where higher-order modulation

Copyright (c) 2010 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org. M. Hansen and O. Winther are with DTU Informatics, Technical University of Denmark, DK-2800 Lyngby, e-mail: {mha,owi}@imm.dtu.dk.

L. P. B. Christensen is with Modem Algorithm Design, Nokia Denmark, Frederikskaj, DK-1790 Copenhagen V., e-mail: lars.christensen@nokia.com.

Manuscript submitted, June 22, 2009; revised February 22, 2010.

schemes are employed. In such scenarios the optimal symbol-by-symbol or sequence detector will often require a very high complexity, due to its exponential growth in complexity as a function of the filter length. Furthermore, in multi user detection the complexity grows further, since the number of users will also influence the complexity exponentially. Thus, suboptimal schemes, such as delayed decision feedback, [4], or reduced-state sequence estimation, [5], will often be applied in such systems instead [6]. However, in order to ensure acceptable performance of these schemes, both the minimum-phase filter and the associated all-pass filter are usually needed [4], since the minimum-phase filter provides the highest possible energy concentration in the beginning of the filter impulse response [2].

The scope of this paper is to show a new method of computing both the minimum-phase filter and the associated all-pass filter using the QL-factorization. This insight provides an alternative approach for computing the minimum-phase filter in a numerically stable way, due to the Householder transformation [7], [8]. Furthermore, as shown in [9], fast low-complexity algorithms can be exploited when computing the QL-factorization of the filtering matrix [10]–[13], if one is willing to sacrifice numerical precision in favor of reduced complexity. Thus, the minimum-phase and all-pass filters can be computed efficiently using this method (see [9] for a more detailed description of this).

The rest of the paper is organized as follows; In Section II we present the general system setup, including the system model, while Section III describes some of the existing methods for computing the minimum-phase filter, and Section IV presents an overall treatment of why the QL-factorization provides the minimum-phase filter. Section V contains an elaborate proof of this as well as an analysis of the convergence rate. In Section VI some simulation results are shown, and some concluding remarks are given in Section VII.

Throughout the paper bold lowercase letters (e.g. \mathbf{x}) denote column vectors, while bold uppercase letters denote matrices (e.g. \mathbf{H}). The matrix transpose is denoted $(\cdot)^T$, while $(\cdot)^H$ is the Hermitian matrix transpose, and the complex conjugate of a complex number is represented by $(\cdot)^*$.

II. SYSTEM MODEL

We consider a time-invariant Multiple-Input Multiple-Output (MIMO) system with a Finite Impulse Response (FIR) length L . The output signal $\mathbf{y}_j \in \mathbb{C}^{N_R}$ at time index j can be expressed as

$$\mathbf{y}_j = \sum_{l=0}^{L-1} \mathbf{H}_l \mathbf{x}_{j-l} + \mathbf{v}_j, \quad (1)$$

where $\mathbf{x}_j \in \mathbb{C}^{N_T}$ is the input signal at time $j = \{1, 2, \dots, J\}$, and $\mathbf{v}_j \in \mathbb{C}^{N_R}$ represents the noise term, $\mathbf{v}_j \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I})$. J is the length of the input sequence and N_R and N_T denote the size of input and

output vectors, respectively (in communications also called receive and transmit dimensions). We assume that $N_R \geq N_T$ which implies that the matrix $\mathbf{H}_l \in \mathbb{C}^{N_R \times N_T}$ denoting the l th tap in the impulse response, will be either a “tall-thin” or a square matrix. Using matrix notation, the system model in (1) can be formulated as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{v} , \quad (2)$$

where $\mathbf{y} = [\mathbf{y}_1^T, \mathbf{y}_2^T, \dots, \mathbf{y}_{J+L-1}^T]^T$ and $\mathbf{x} = [\mathbf{x}_1^T, \mathbf{x}_2^T, \dots, \mathbf{x}_J^T]^T$. To ease the notation let $M \triangleq N_R(J+L-1)$ and $N \triangleq JN_T$, leading to $\mathbf{y} \in \mathbb{C}^M$ and $\mathbf{x} \in \mathbb{C}^N$. Due to the time-invariant property of the filter, $\mathbf{H} \in \mathbb{C}^{M \times N}$ will be a block-banded block Toeplitz convolution matrix (also referred to as the filtering matrix), having the form

$$\mathbf{H} \triangleq \begin{bmatrix} \mathbf{H}_0 & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{H}_{L-1} & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \ddots & \ddots & \mathbf{H}_0 \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{H}_{L-1} \end{bmatrix} .$$

The finite-length system above can be described in polynomial form when we let the system size $J \rightarrow \infty$. The connection between the finite- vs. the infinite-length system is, among others, treated in [14], and the z-transform of the equivalent infinite-length filter impulse response, is given as

$$\mathbf{H}(z) = \sum_{l=0}^{L-1} \mathbf{H}_l z^{-l} , \quad (3)$$

which is a useful representation in the analysis of the filter characteristics.

By QL-factorizing the filtering matrix, $\mathbf{H} = \mathbf{Q}\mathbf{L}$ in (2) and multiplying by \mathbf{Q}^H with (2), we get a new equivalent system equation

$$\tilde{\mathbf{y}} \triangleq \mathbf{Q}^H \mathbf{y} = \mathbf{L}\mathbf{x} + \tilde{\mathbf{v}} , \quad (4)$$

where we have used the fact that \mathbf{Q} is a unitary matrix. Importantly, it also follows from unitarity that the noise statistic is unchanged (under the assumption of Gaussian noise). Fig. 1 illustrates the system model given in (4).

III. THEORY ON THE MINIMUM-PHASE FILTER

As mentioned in the introduction, the minimum-phase filter has been studied intensively over the years due to its broad applicability, and there are various methods for computing the filter. In [15], [16] and the

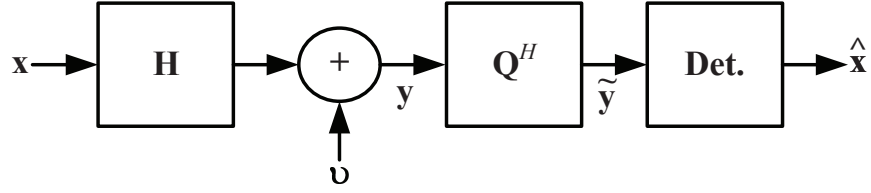


Figure 1. System model with prefilter and detection stage included.

references therein a thorough treatment of several methods for spectral factorization can be found. One classical way of obtaining the minimum-phase filter is by using the root-method of spectral factorization, in which the roots of (3) - which for $N_R = N_T$ satisfy $\det(\mathbf{H}(z)) = 0$ - located outside the unit circle are reflected inside to the conjugate reciprocal location [1], [2], [17]. This simple method however has its limitations, particularly in the case of vector observations (i.e. MIMO systems), since besides the roots we also need to know the direction of the vector associated with that root [3]. Some methods for solving the problem in this case have been described in, among others [18]–[21], but these methods have the disadvantage of being mathematically rather complicated and, furthermore, can suffer from numerical instabilities [3, p. 206]. Thus, one might prefer to solve a Discrete-time Algebraic Riccati Equation (DARE) instead, which is a numerical stable method, having the particularly advantageous property that it can easily be extended to the vector case [15]. In the following we briefly describe how the roots can be determined, which we will be using in the analysis of the convergence rate.

A. The root-method of spectral factorization

Let us for a moment assume that we are only interested in determining the roots of $\mathbf{H}(z)$ in (3). In a MIMO system where $N_T = N_R$, the roots can be obtained by finding the z -values where $\det(\mathbf{H}(z)) = 0$, [22], leading to a matrix polynomial in the scalar variable z . This type of matrix polynomial is normally called a lambda-matrix [23], [24] and the number of roots in such a polynomial is $\min\{N_T, N_R\}(L - 1)$. In [23], it is shown that the roots can be obtained by determining the eigenvalues of the block-companion matrix, \mathbf{C} , of the associated *monic* polynomial, which can be obtained by $\tilde{\mathbf{H}}(z) \triangleq (\mathbf{H}_{L-1})^{-1} \mathbf{H}(z)$, where

we have assumed that \mathbf{H}_{L-1} is invertible. Thus, we get the following block-companion matrix

$$\mathbf{C} \triangleq \begin{bmatrix} \mathbf{0} & \cdots & \mathbf{0} & -\tilde{\mathbf{H}}_0 \\ \mathbf{I} & \ddots & \vdots & -\tilde{\mathbf{H}}_1 \\ \mathbf{0} & \ddots & \mathbf{0} & \vdots \\ \vdots & \ddots & \mathbf{I} & -\tilde{\mathbf{H}}_{L-2} \end{bmatrix}, \quad (5)$$

where $\tilde{\mathbf{H}}_k \triangleq (\mathbf{H}_{L-1})^{-1} \mathbf{H}_k$. Since the method proposed in [23], [24] assumes that all \mathbf{H}_l terms are square matrices, we cannot directly handle the case where $N_T \neq N_R$ and, therefore, we need to modify the problem. If $N_R > N_T$ we can instead introduce $\mathbf{S} = \mathbf{H}^H \mathbf{H}$ and find the roots of the lambda-matrix based on $\mathbf{S}(z)$,

$$\begin{aligned} \mathbf{S}(z) &= \mathbf{H}^H(z^{-*}) \mathbf{H}(z) \\ &= \sum_{l=0}^{L-1} \mathbf{S}_l z^{-l} + \sum_{k=1}^{L-1} \mathbf{S}_k^H z^k, \end{aligned} \quad (6)$$

giving the roots both inside and outside the unit circle (from the minimum- and maximum-phase filter, respectively). This does not however solve the problem of finding the zero directions in the lambda-matrix, and we will therefore also address an alternative way of computing the spectral factor.

B. The DARE Method

As mentioned in the previous subsection, the DARE method has the convenient property that it is straight forward to extend from the Single-Input Single-Output (SISO) case to the MIMO case. Furthermore, the method relates to results from Kalman filtering theory and, therefore, many of the properties of this method have been extensively studied, among others its convergence properties, [3]. The DARE method considered in this paper, solves the Riccati equation using the iterative procedure described in [16],¹ and once the stabilizing solution has been obtained, the filter coefficients are computed as described in [25]. The complexity of computing the minimum-phase filter of a length L SISO system using the DARE method is

$$\mathcal{O}_{min,DARE} = k \left(\frac{3}{2} L^2 - \frac{1}{2} L + 2 \right) + 2L \quad (7)$$

operations, where k denotes the number of iterations used for computing the filter.²

¹In [16] the procedure is referred to as *the method of doubling*.

²We here define an operation as a complex Multiply-Accumulate (MAC) instruction.

IV. CONNECTION BETWEEN THE MINIMUM-PHASE FILTER AND THE QL-FACTORIZATION

As mentioned earlier, it is well-known that the minimum-phase filter can be computed in several ways, and recently it has been realized that the minimum-phase filter and the associated all-pass filter can be obtained by QL-factorizing the filtering matrix \mathbf{H} , [26], [27], such that

$$\mathbf{H} = \mathbf{Q}\tilde{\mathbf{L}} = \mathbf{Q} \begin{bmatrix} \mathbf{0}_{(M-N) \times N} \\ \mathbf{L} \end{bmatrix}, \quad (8)$$

where $M \geq N$, and we require that the $N \times N$ lower triangular matrix, \mathbf{L} , corresponds to the Cholesky factor of $\mathbf{H}^H \mathbf{H}$, implying that \mathbf{L} is positive definite and thus contains real-valued positive diagonal elements (assuming that $\text{rank}(\mathbf{H}) = N$). In Section V we prove the connection between the minimum-phase filter and the QL-factorization in a more formal way compared to the argument presented in [27], but we would first like to repeat the intuitive argument in order to clarify why we can obtain minimum-phase and all-pass filters using the QL-factorization.

When we QL-factorize the time-invariant block-banded block Toeplitz matrix, each block-row in \mathbf{L} will be a shifted version of each other for $N \rightarrow \infty$, where each block-row is given by the spectral factorization, [28]. Likewise, the $M \times M$ unitary matrix, \mathbf{Q} , will be the matrix equivalent to the all-pass filter, where again each block-column of \mathbf{Q} will be a shifted version of each other (for $N \rightarrow \infty$). Furthermore, it can be seen that each of these block-columns will correspond to the finite dimensional analog of the all-pass filter associated with the minimum-phase filter. In the finite length case, each block-row of \mathbf{L} (block-column of \mathbf{Q}) will not be exactly the same, but as we will show later in the paper, the values in each of these will converge toward the true minimum-phase filter as a function of the block-row number.³ The block-columns of \mathbf{Q} will similarly converge toward the associated all-pass filter.

As shown in [9] the complexity of obtaining the minimum-phase filter of a SISO system using fast QL-factorization is

$$\mathcal{O}_{min,QL} = (k-1)(5L+7) + (L-1)L/2 + 4L \quad (9)$$

operations and $k+1$ square root operations, while the all-pass filter requires

$$\begin{aligned} \mathcal{O}_{ap,QL} = & \sum_{n=0}^{\min\{(L_{ap}-1);(k-1)\}} \min\{(L+n); L_{ap}\}(n+1) \\ & + \max\{0; (k-L_{ap})\} L_{ap}(L_{ap}+1) \end{aligned} \quad (10)$$

³Strictly speaking the elements in the block-row of \mathbf{L} converge toward the minimum-phase filter from the bottom up, since the Householder transformation computes the elements in the lower triangular matrix from the bottom (when we perform a QL-factorization instead of a QR-factorization).

operations, assuming that $k \geq L_{ap}$. Here L_{ap} denotes the length of the all-pass filter. From (7) and (9) we see that the fast QL-factorization has a computational advantage over the DARE method described in Section III-B, which can also be seen from Table I in Section VI where a comparison of the complexity of the methods is given.

It should be noted that in the case where we have a time-variant filter, the block-rows in the lower triangular matrix will in some sense represent an analog to the “instantaneous” minimum-phase filter and, likewise, the block-columns of unitary matrix will represent the associated “instantaneous” all-pass filter. It should also be noted that if we perform a QR-factorization of the filtering matrix instead of a QL-factorization, we will get the *maximum-phase* filter.

A. The Householder Transformation

In our analysis of the convergence toward the minimum-phase filter and the all-pass filter we use the Householder transformation to compute the QL-factorization. Therefore, we first briefly describe the steps of this transformation. The reason for choosing this transformation is its advantageous numerical stability to roundoff effects. For a more thorough treatment of the transformation and its numerical properties the reader is referred to [7]. In most textbooks, the Householder algorithm is only described for real numbers and since this transformation plays a crucial role in our treatment of the convergence rate in Section V, we here illustrate a complex version of the transformation. The Householder transformation (for QL-factorization) of a matrix $\mathbf{B} \in \mathbb{C}^{M \times N}$ works as illustrated in Algorithm 1, where \mathbf{e}_k denotes the unit vector with 1 in the k th position, and where we have defined the unitary matrices $\mathbf{U}_k \in \mathbb{C}^{M \times M}$ and $\hat{\mathbf{U}}_k \in \mathbb{C}^{(M-k+1) \times (M-k+1)}$ as

$$\mathbf{U}_k \triangleq \begin{bmatrix} \hat{\mathbf{U}}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{k-1} \end{bmatrix}.$$

V. CONVERGENCE RATE

In this section we examine the convergence properties of the rows and columns the QL-factorization toward the minimum-phase and all-pass filters. In order to simplify this analysis, we first consider the simplest possible case, which is for the SISO case with a filter length of $L = 2$. We will then extend this result to the more general one.

Algorithm 1 Householder Transformation for QL-fact.

```

1: Input: Matrix  $\mathbf{B}$ 
2:  $\hat{\mathbf{B}} \leftarrow \mathbf{B}$ 
3: for  $k = 1$  to  $\min\{M, N\}$  do
4:   {Pick out the last column vector of  $\hat{\mathbf{B}}$ }
5:    $\hat{\mathbf{b}} = \hat{\mathbf{B}}_{:,end}$ 
6:    $\tilde{k} = M - k + 1$ 
7:   {Do Householder reflection of  $\hat{\mathbf{b}}$  (line 8 to 12)}
8:    $\alpha = \|\hat{\mathbf{b}}\|$ 
9:    $\tilde{\alpha} = e^{i\angle \hat{\mathbf{b}}_{\tilde{k}}} \alpha$ 
10:   $\mathbf{v} = \hat{\mathbf{b}} + \tilde{\alpha} \mathbf{e}_{\tilde{k}}$ 
11:   $\hat{\mathbf{U}}_k = e^{-i\angle \hat{\mathbf{b}}_{\tilde{k}}} \left( \frac{2\mathbf{v}\mathbf{v}^H}{\|\mathbf{v}\|^2} - \mathbf{I} \right)$ 
12:   $\tilde{\mathbf{B}} = \hat{\mathbf{U}}_k \hat{\mathbf{B}}$ 
13:  {Remove last row and last column of  $\tilde{\mathbf{B}}$ }
14:   $\hat{\mathbf{B}} \leftarrow \tilde{\mathbf{B}}_{1:(end-1), 1:(end-1)}$ 
15:  {Repeat for new  $\hat{\mathbf{B}}$ }
16:   $\hat{\mathbf{B}} \leftarrow \hat{\mathbf{B}}$ 
17: end for

```

After $K = \min\{M, N\}$ iterations we have;

$$\mathbf{L} = \mathbf{U}_K \dots \mathbf{U}_2 \mathbf{U}_1 \mathbf{B}$$

$$\mathbf{Q} = \mathbf{U}_1^H \mathbf{U}_2^H \dots \mathbf{U}_K^H$$

A. SISO system with filter length $L = 2$

Any SISO filtering matrix of a length $L = 2$ system can be formulated as

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ a & 1 & \ddots & \vdots \\ 0 & a & \ddots & 0 \\ \vdots & \ddots & \ddots & 1 \\ 0 & \cdots & 0 & a \end{bmatrix},$$

where we have normalized the impulse response such that $a \triangleq h_1/h_0 \neq 0$, leading to $H(z) = 1 + az^{-1}$ for the equivalent infinite-length filter impulse response. In this case it is trivial to compute the minimum-

phase solution using the root-method

$$z_{mp} = \begin{cases} -a & \text{if } |a| \leq 1 \\ -1/a^* & \text{else} \end{cases}, \quad (11)$$

where z_{mp} represents the minimum-phase root. Since $H(z) = H_{ap}(z)H_{mp}(z)$ we have

$$H_{ap}(z) = \begin{cases} 1 & \text{if } |a| \leq 1 \\ \frac{z^{-1} + \frac{1}{a}}{1 + \frac{1}{a^*} z^{-1}} & \text{else} \end{cases}, \quad (12)$$

where $H_{mp}(z)$ and $H_{ap}(z)$ represent the z -transformed minimum-phase and all-pass filters, respectively.⁴

By QL-factorizing the filtering matrix we get

$$\mathbf{L} = \begin{bmatrix} \alpha_N & 0 & \cdots & 0 \\ \beta_{N-1} & \alpha_{N-1} & \ddots & \vdots \\ 0 & \beta_{N-2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \alpha_2 & 0 \\ 0 & \cdots & 0 & \beta_1 & \alpha_1 \end{bmatrix}, \quad (13)$$

where we are interested in determining the α and β values. For notational brevity we introduce $\gamma_k \triangleq \hat{b}_{\tilde{k}}$, where $\hat{b}_{\tilde{k}}$ is defined in Algorithm 1 as the last element in vector $\hat{\mathbf{b}}$, which is being reflected in the k th iteration. Since $\hat{\mathbf{b}}_1 = [0, \dots, 0, 1, a]^T$ we have $\gamma_1 = a$ for the first Householder reflection (also referred to as iteration $k = 1$). Based on the input vector we see from line 8 in Algorithm 1 that $\alpha_1 = \sqrt{1 + |\gamma_1|^2}$ and from lines 9-10 we get $\mathbf{v}_1 = [0, \dots, 0, 1, \gamma_1 + \tilde{\alpha}_1]^T$. Lines 11 and 12 in Algorithm 1 lead to the following expression for the β ,

$$\beta_1 = \frac{2a e^{-i\angle\gamma_1} (\gamma_1 + e^{i\angle\gamma_1} \alpha_1)}{1 + |\gamma_1 + e^{i\angle\gamma_1} \alpha_1|^2} \quad (14a)$$

$$= \frac{2a (|\gamma_1| + \alpha_1)}{1 + (|\gamma_1| + \alpha_1)^2}, \quad (14b)$$

where in (14b) we have used $\gamma_1 = e^{i\angle\gamma_1} |\gamma_1|$. After the first Householder reflection we have

$$\mathbf{U}_1 \mathbf{H} = \begin{bmatrix} \ddots & 0 & \cdots & 0 \\ \ddots & 1 & \ddots & \vdots \\ \ddots & a & 1 & 0 \\ \vdots & 0 & \gamma_2 & 0 \\ \cdots & 0 & \beta_1 & \alpha_1 \end{bmatrix},$$

⁴In order to ensure that the magnitude response of the all-pass filter will always be one, we have normalized the minimum-phase filter such that $H_{mp}(z) = a(1 + 1/a^* z^{-1})$ whenever a root is reflected inside the unit circle (i.e. when $|a| > 1$).

since there are only two non-zero elements in the columns of \mathbf{H} . In the next iteration we will have $\hat{\mathbf{b}}_1 = [0, \dots, 0, 1, \gamma_2]^T$, and by examining the update steps in the Householder reflection carefully, it becomes clear that the value of γ_{k+1} can be expressed as a function of γ_k , leading to a recursive update given as

$$\gamma_{k+1} = a \left(1 - \frac{2}{1 + |\gamma_k + e^{i\angle\gamma_k} \alpha_k|^2} \right) \quad (15a)$$

$$= a \left(1 - \frac{2}{1 + \left| |\gamma_k| + \sqrt{1 + |\gamma_k|^2} \right|^2} \right) \quad (15b)$$

$$= \frac{a |\gamma_k|}{\sqrt{1 + |\gamma_k|^2}}. \quad (15c)$$

Likewise, the general expression for the α 's and β 's will be

$$\alpha_k = \sqrt{1 + |\gamma_k|^2} \quad (16)$$

$$\beta_k = \frac{2a (|\gamma_k| + \alpha_k)}{1 + |\gamma_k| + \alpha_k} \quad (17)$$

From (16) we can verify that the α values will be positive and real-valued, which is exactly what is required from the QL-factorization. From (15) and (17) we also see the interesting property that all the values of the γ_k 's and the β_k 's will always have the same angle in the complex plane, determined by $\angle\beta_k = \angle\gamma_k = \angle a$. This implies that the convergence of the β_k 's to the true minimum-phase solution for each iteration takes place in the same direction in the complex plane.

Lemma V.1 (Recursive computation of α_k and β_k). *In a time-invariant SISO system with $L = 2$, the coefficients in \mathbf{L} obtained by the Householder transformation can be determined as*

$$\alpha_k = \sqrt{1 + |\gamma_k|^2}$$

$$\beta_k = \frac{2a (|\gamma_k| + \alpha_k)}{1 + |\gamma_k| + \alpha_k}$$

where

$$\gamma_{k+1} = \frac{a |\gamma_k|}{\sqrt{1 + |\gamma_k|^2}}.$$

Proof: Given above. ■

As shown in Appendix A the recursive expression for γ_k given in (15), can be rewritten as

$$\gamma_k = e^{i\angle a} \sqrt{\frac{|a|^2 - 1}{1 - |a|^{-2k}}}. \quad (18)$$

Now in order to show that the values of α_k and β_k match the minimum-phase filter, we need to determine the fixed-point solutions for the parameter γ_k in (15), such that

$$\gamma_{fix} = f(\gamma_{fix}), \text{ where } f(x) = \frac{a|x|}{\sqrt{1+|x|^2}}.$$

As shown in the lemma below, there are two fixed-points.

Lemma V.2 (Fixed-points for γ). *In a time-invariant SISO system with $L = 2$, the fixed-point solutions for γ will be*

$$\gamma_{fix} = \begin{cases} 0 & \text{if } |a| \leq 1 \\ e^{i\angle a} \sqrt{|a|^2 - 1} & \text{else} \end{cases}.$$

Proof: See Appendix B for a detailed proof. ■

Based on these fixed-points for γ we have

$$\alpha_{fix} = \begin{cases} 1 & \text{if } |a| \leq 1 \\ |a| & \text{else} \end{cases} \quad (19a)$$

$$\beta_{fix} = \begin{cases} a & \text{if } |a| \leq 1 \\ e^{i\angle a} & \text{else} \end{cases}. \quad (19b)$$

Thus, the root of \mathbf{L} obtained using the QL-factorization, will be $z_L = -\beta_{fix}/\alpha_{fix}$

$$\begin{aligned} z_L &= \begin{cases} -a & \text{if } |a| \leq 1 \\ -e^{i\angle a}/|a| & \text{else} \end{cases} \\ &= \begin{cases} -a & \text{if } |a| \leq 1 \\ -1/a^* & \text{else} \end{cases}, \end{aligned} \quad (20)$$

which corresponds to the result given in (11), obtained by the traditional root-method of spectral factorization. Likewise, the unitary matrix will converge to the Infinite Impulse Response (IIR) all-pass filter given in (12). In order to ensure that we do in fact get the minimum-phase solution, we also need to prove that the recursive expression for γ_k converges to the fixed-points. In Appendix C it has been proved that this is indeed the case. Thus, it can be concluded that in the SISO case with a filter length of $L = 2$, the elements in the rows of \mathbf{L} converge to the minimum-phase filter.⁵

⁵This is no surprise, since it has already been shown in [15], [28] that the lower triangular matrix provides the spectral factor.

In the following we examine the convergence rate to the fixed-point solutions, which can be determined based on the expression for γ_k given in (18). In order to compute the convergence rate we introduce

$$\gamma_k = \gamma_{fix} + \Delta\gamma_k, \quad (21)$$

where $\Delta\gamma_k$ represents the deviation of γ_k from the fixed-point solution. To upper bound the convergence we treat the cases of $|a| \leq 1$ and $|a| > 1$ separately.

1) *The $|a| \leq 1$ case:* From (18) we get that

$$|\Delta\gamma_k| = |\gamma_k - \gamma_{fix}| = |a|^k \sqrt{\frac{|a|^2 - 1}{|a|^{2k} - 1}} \quad (22a)$$

$$\leq |a|^k \sqrt{\frac{|a|^2 - 1}{|a|^2 - 1}} = |a|^k \quad \text{for } \forall k \geq 1. \quad (22b)$$

2) *The $|a| > 1$ case:* When $|a| > 1$ the fixed-point is $|\gamma_{fix}| = \sqrt{|a|^2 - 1}$ and from Lemma C.1 we know that $|\gamma_k| \geq |\gamma_{fix}|$. As mentioned in Appendix C all of the terms which are compared have the same argument and, therefore, we can simply ignore the angle and only consider the case where the terms are real and positive. We then get

$$|\Delta\gamma_k| = |\gamma_k| - |\gamma_{fix}| \quad (23a)$$

$$= \sqrt{|a|^2 - 1} \left(\frac{1}{\sqrt{1 - |a|^{-2k}}} - 1 \right) \quad (23b)$$

$$\leq \sqrt{|a|^2 - 1} \left(\frac{1}{1 - |a|^{-2k}} - 1 \right) \quad (23c)$$

$$\leq |a|^{-2k} \frac{\sqrt{|a|^2 - 1}}{1 - |a|^{-2}} = |a|^{-2k} \frac{|a|^2}{\sqrt{|a|^2 - 1}}. \quad (23d)$$

Thus, we have the following lemma which upper bounds the convergence rate.

Lemma V.3 (Upper bound on the convergence rate of γ). *In a time-invariant SISO system with $L = 2$, the convergence rate of γ_k can be upper bounded by*

$$|\Delta\gamma_k| \leq |\Delta\tilde{\gamma}_k| = \begin{cases} e^{k \ln(|a|)} & \text{if } |a| \leq 1 \\ \frac{|a|^2}{\sqrt{|a|^2 - 1}} e^{2k \ln(1/|a|)} & \text{else} \end{cases}.$$

From Lemma V.3 we see the interesting property that the convergence rate is exponential and is determined by $|a|$. In other words, the convergence rate to the fixed-point solution is governed by the localization of the root in the complex plane. In the case where we have a root which is close to the unit

circle, we will have slow convergence to the minimum-phase solution. In Fig. 2(a) the convergences of $\Delta\gamma_k$ and $\Delta\tilde{\gamma}_k$ have been shown as a function of the number of iterations for a $L = 2$ SISO system in the case where the root is $z = \{-0.3, -0.6, -0.9\}$, respectively. From the figure it is clearly seen that the distance between the root and the unit circle has a significant influence on the convergence rate and, furthermore, we see that the upper bound becomes tighter as the distance grows. It is also relevant to examine how the deviation $\Delta\gamma_k$ affects the value of the root. Therefore, we introduce $z_{L,k} \triangleq -\beta_k/\alpha_k$, which represents the root obtained from \mathbf{L} in the k th iteration. Likewise, we have $\tilde{z}_{L,k} \triangleq -\tilde{\beta}_k/\tilde{\alpha}_k$, where the approximated values of α_k and β_k have been obtained using $\Delta\tilde{\gamma}_k$.

In Fig. 2(b) the deviations from the true minimum-phase root have been plotted, where we have defined $\Delta z_k \triangleq z_{mp} - z_{L,k}$ and $\Delta\tilde{z}_k \triangleq z_{mp} - \tilde{z}_{L,k}$. From the figure, we see that the deviation $\Delta\gamma_k$ is significantly larger than the deviation in the root value Δz_k .

B. SISO system with filter length $L > 2$

In the case where we have a filter length of $L > 2$, the deviations of the recursions for the Householder transformation become much more complicated, since the vector \mathbf{b} in Algorithm 1 will now have L non-zero elements. Thus, it will no longer be the simple scalar recursion for γ_k but instead a $(L-1) \times (L-1)$ matrix recursion and, furthermore, due to the multiple roots there will also be multiple fixed points. However, we can generalize the result obtained for the $L = 2$ SISO system by factorizing the filtering matrix into $(L-1)$ products of $L = 2$ filtering matrices,⁶ such that

$$\mathbf{H} = \mathbf{H}_2^{(L-1)} \mathbf{H}_2^{(L-2)} \dots \mathbf{H}_2^{(1)}, \quad (24)$$

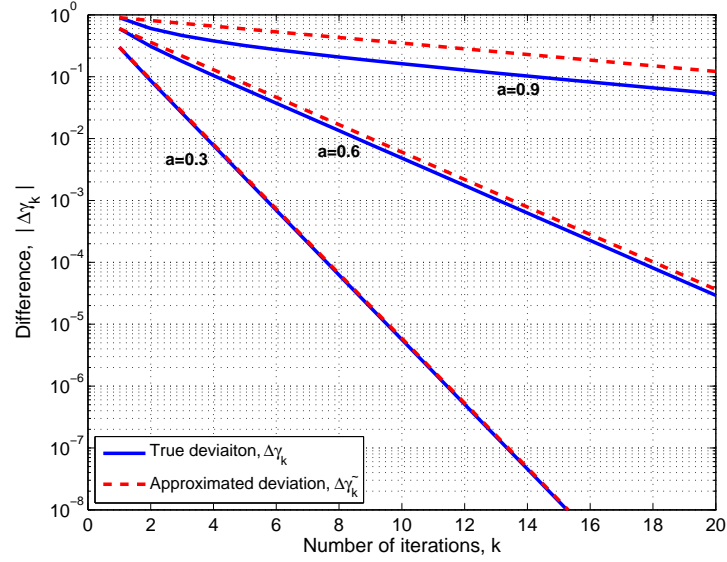
here $\mathbf{H}_2^{(l)}$ is the filtering matrix of the l th length two filter, where the z -transform of the equivalent infinite-length filter impulse response is given as $H_2^{(l)}(z) \triangleq 1 + a_l z^{-1}$. The factorization makes it possible to perform a QL-factorization on each of the $(L-1)$ terms in (24), which gives

$$\mathbf{H}_2^{(l)} = \mathbf{Q}_2^{(l)} \mathbf{L}_2^{(l)}. \quad (25)$$

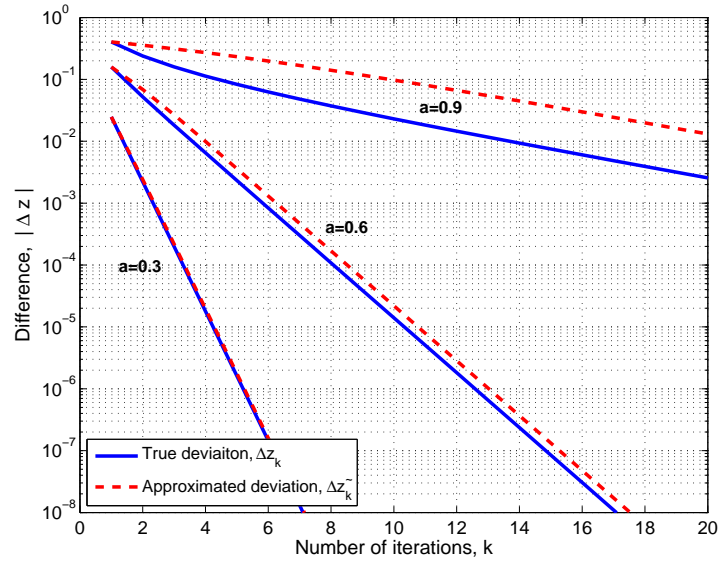
where the convergence rate of each of the $(L-1)$ terms is given in Subsection V-A. By inserting (25) into (24) we get

$$\mathbf{H} = \mathbf{Q}\mathbf{L} = \mathbf{Q}_2^{(L-1)} \mathbf{L}_2^{(L-1)} \mathbf{Q}_2^{(L-2)} \mathbf{L}_2^{(L-2)} \dots \mathbf{Q}_2^{(1)} \mathbf{L}_2^{(1)}. \quad (26)$$

⁶It should be noted that the size of $\mathbf{H}_2^{(l)}$ decreases by one (both column- and row-wise) as l decreases by one, in order to enable the factorization.



(a) True and approximated value of $\Delta\gamma_k$.



(b) True and approximated value of root-deviation Δz_k .

Figure 2. Example of deviations of $\Delta\gamma_k$ and Δz_k in a SISO system, with length $L = 2$ and root at $a = \{0.3, 0.6, 0.9\}$, respectively

We would like to reorder the terms on the RHS of (26) such that all $\mathbf{Q}_2^{(l)}$ terms are grouped together followed by all the $\mathbf{L}_2^{(l)}$ terms, i.e.

$$\mathbf{H} \cong \underbrace{\mathbf{Q}_2^{(L-1)} \mathbf{Q}_2^{(L-2)} \dots \mathbf{Q}_2^{(1)}}_{\mathbf{Q}} \underbrace{\mathbf{L}_2^{(L-1)} \mathbf{L}_2^{(L-2)} \dots \mathbf{L}_2^{(1)}}_{\mathbf{L}}, \quad (27)$$

where the equality holds when the system size $N \rightarrow \infty$. The reason that it is possible to rearrange the terms when the system size goes to infinity is due to fact that $\mathbf{L}_2^{(l)}$ and $\mathbf{Q}_2^{(l)}$ asymptotically become circulant matrices [29], and thereby, we can use the commutative property of circulant matrices [29]. Conceptually it is fairly easy to see why $\mathbf{L}_2^{(l)}$ asymptotically becomes circulant, since it is a banded matrix, but this might not be as obvious for the all-pass filtering matrix, which represents an IIR filter. However, it has been proved in [30] that the IIR filter has an exponential decay, which implies that, in the limit where the system size tends to infinity, the IIR filter becomes a Toeplitz matrix. In [29] it is proved that general Toeplitz matrices containing absolutely summable elements (also referred to as *Wiener Class Toeplitz Matrices*) asymptotically converge to circulant matrices too. Thus, in the limit $N \rightarrow \infty$ both matrices become circulant and, therefore, we know that the lower triangular matrix \mathbf{L} converge to the minimum-phase filter for SISO systems of arbitrary length. Due to the unique factorization of $\mathbf{H} = \mathbf{Q}\mathbf{L}$ (where we require that the elements on the diagonal of \mathbf{L} are real-valued and positive), \mathbf{Q} must be the matrix version of the all-pass filter associated with the minimum-phase filter, since it is the only unitary matrix which links \mathbf{L} with \mathbf{H} .

Based on the expression in (27) it is possible to approximate the convergence rate in a SISO system of arbitrary length, by examining the deviations in the approximated root values $\Delta \tilde{z}_k^{(l)} \triangleq z_{mp}^{(l)} - \tilde{z}_{L,k}^{(l)}$, where $z_{mp}^{(l)}$ represents the l th root of the true minimum-phase filter and $\tilde{z}_{L,k}^{(l)} \triangleq -\tilde{\beta}_k^{(l)} / \tilde{\alpha}_k^{(l)}$ is the approximated value of the l th root based on the upper bound given in Lemma V.3. Thus, in the z -domain the difference between the true minimum-phase filter and the filter obtained based on $\tilde{z}_{L,k}^{(l)}$ becomes

$$\Delta H(z) \triangleq H_{mp}(z) - \tilde{L}_k(z) \quad (28a)$$

$$\approx z^{-(L-1)} \left[\prod_{l=1}^{L-1} \left(z - z_{mp}^{(l)} \right) - \prod_{l=1}^{L-1} \left(z - \tilde{z}_{L,k}^{(l)} \right) \right], \quad (28b)$$

where $\tilde{L}_k(z)$ represents the z -transform of the approximate value for the k th row in the lower triangular matrix, \mathbf{L} . In (28) it is only the effect of each deviation in the approximated root value that has been taken into account, but the approximation made in (27) will to some extent affect the convergence rate due to roots only being asymptotically independent. The tightness of the approximate expression for convergence when $L > 2$ has been evaluated empirically in Figs. 4 and 5 given in Section VI. In (28b)

we have normalized the first coefficient and from the equation we can see that the main contribution to the difference between the true minimum-phase filter and the result obtained by the QL-factorization, will asymptotically come from the root which is closest to the unit circle. This observation fits well with what is described in [3, p. 508], where the convergence to the stabilizing solution of the DARE is exponential and determined by the spectral radius.

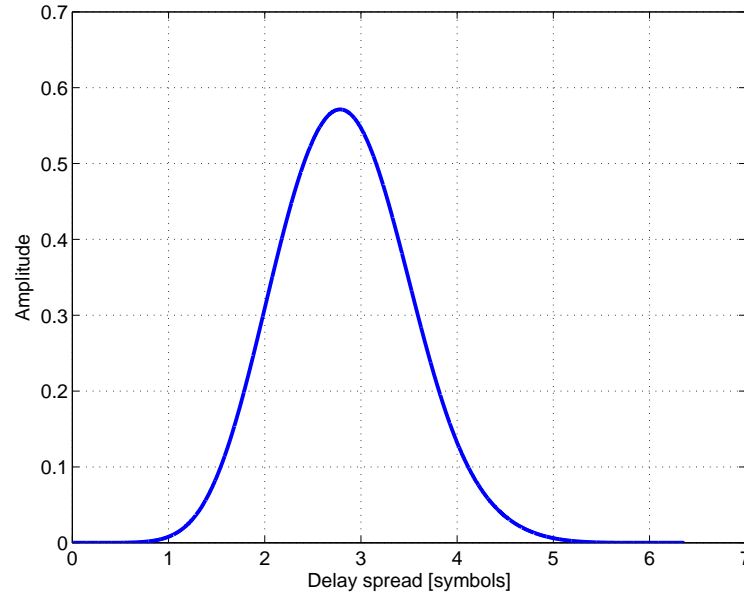
C. MIMO system

In the case where we have a MIMO system, we can first examine the length 2 system $\mathbf{H}(z) = \mathbf{I} + \mathbf{H}_1 z^{-1}$ where $N_T = N_R$. Compared to the SISO system of the same length, the only difference is that the operations now become $N_R \times N_T$ matrix operations instead of scalars. However, since we have $\eta = \min\{N_T, N_R\}(L-1)$ roots, we get 2^η fixed-points, thus it becomes more complicated to analyze even a simple $L = 2$ MIMO system. In the case where we have an arbitrary filter length, the argument presented in Subsection V-B, concerning the SISO system of length $L > 2$, can be repeated here.

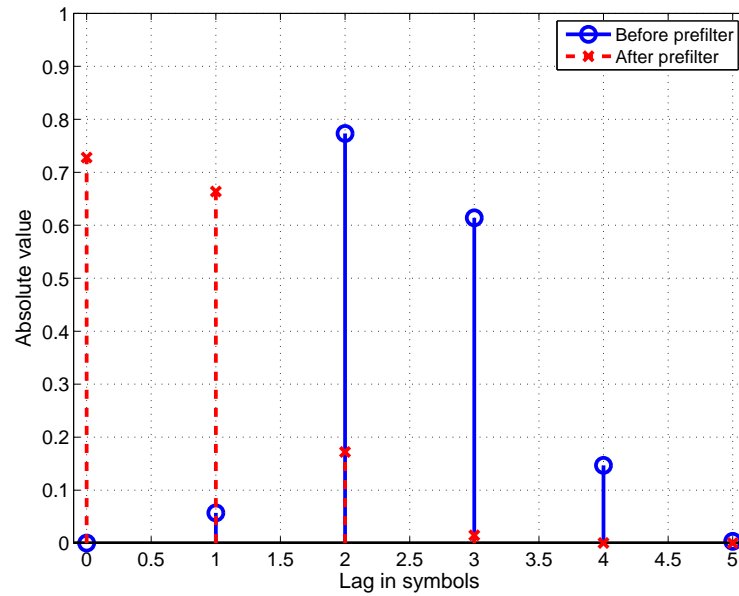
VI. SIMULATION RESULTS

In this section simulation results for both SISO and MIMO systems are presented. For the SISO system we examine two channel scenarios. In the first scenario we have complex Gaussian distributed, $\mathcal{CN}(0, 1)$, filter coefficients. In the second scenario we consider a channel defined in the GSM specifications [31], namely the Typical Urban (TU0) profile. The channel profile, obtained by the convolution of the square root of the power delay profile with the transmit filter response (the so-called C_0 -pulse in [32]), is plotted in Fig. 3, and from the figure it is seen that roughly speaking we need about four symbols to capture the energy of the pulse. If the cardinality of the alphabet is denoted by $|\Omega|$, this implies that a Delayed Decision Feedback Equalizer Sequence Estimation (DDFSE) type of equalizer would require about $|\Omega|^{(4-1)}$ states in the state-space model in order to achieve a performance close to the optimal Maximum Likelihood Sequence Detector (MLSD). The DDFSE equalizer will then have a complexity similar to the MLSD, if no prefiltering is made. In e.g. EGPRS2 [33], this would lead to an unacceptably high decoding complexity, since we can have cardinalities up to $|\Omega| = 32$, giving approximately $3.3 \cdot 10^4$ states. Thus, in such applications it is a great advantage to prefilter the pulse with the all-pass filter to obtain the minimum-phase filter. In [27] we have shown a practical application of this, where we have evaluated the effect of prefiltering in terms of performance of equalizers employing reduced-state sequence estimation techniques.

In order to measure the convergence rate of the filter coefficients, we compute the relative difference



(a) The ensemble average of the pulse shape of the TU profile (including the transmit pulse shaping).



(b) The absolute value of filter coefficients for one realization of the TU profile with and without minimum-phase prefiltering.

Figure 3. Delay profile and single realization thereof for the Typical Urban (TU) channel.

between two overall filtering impulse response matrices, $\mathcal{H}_{a,k}$ and $\mathcal{H}_{b,k}$, at the iteration number k , as

$$d(\mathcal{H}_{a,k}; \mathcal{H}_{b,k}) \triangleq \frac{\|\mathcal{H}_{a,k} - \mathcal{H}_{b,k}\|_2}{\|\mathcal{H}_{a,k}\|_2}. \quad (29)$$

We define \mathcal{H}_{mp} as the impulse response of the true minimum-phase filter, and $\mathcal{H}_{L,k}$ represents the impulse response obtained from \mathbf{L} (at iteration k). To measure how well the estimated all-pass filter, $\mathcal{H}_{Q,k}$, matches the estimated minimum-phase filter $\mathcal{H}_{L,k}$, we filter the original impulse response \mathcal{H} with $(\mathcal{H}_{Q,k})^H$, which gives us the output $\mathcal{H}_{\hat{L},k}$.

In all the simulations presented below, we have made 10,000 realizations of the examined channel profile, and computed the minimum-phase and the all-pass filter for each realization. The filter length of the all-pass filter is set to $L_{ap} = 64$ in the simulations. Based on the results obtained from the 10,000 filter realizations, we have computed the mean and median value of the relative errors, $d(\mathcal{H}_{mp}; \mathcal{H}_{L,k})$ and $d(\mathcal{H}_{L,k}; \mathcal{H}_{\hat{L},k})$.

The results for the Gaussian filter coefficients with uniform power in the delay domain are shown in Fig. 4, where we see that the rows in \mathbf{L} converge to the true minimum-phase filter as a function of the iteration number (i.e. the row number).⁷ From the figure we observe that the median value of $d(\mathcal{H}_{mp}; \mathcal{H}_{L,k})$ converges exponentially to zero and that the median difference is about 10^{-8} after 140 iterations. The convergence of the average difference is considerably slower, due to the instances where a channel realization has zeros very close to the unit circle, which will lead to a slow convergence. Thus, these cases tend to bias the estimate of average convergence rate. This is indeed what can be observed from the estimated probability density function (pdf) of $d(\mathcal{H}_{mp}; \mathcal{H}_{L,k})$. Likewise, the mean of $d(\mathcal{H}_{L,k}; \mathcal{H}_{\hat{L},k})$ seems to be biased, which (besides the effect described above) is also due to the truncation of the IIR all-pass filter. Both the mean and median value of the approximated convergence $d(\mathcal{H}_{mp}; \mathcal{H}_{\hat{L},k})$ have also been plotted, and from (29) it is seen that this term represents the energy of the approximated deviation obtained in (28) normalized with respect to the energy of the minimum-phase filter. From the figure it can be seen that the trend of the true and approximated deviation behaves similarly. As a reference we have also included the relative deviation between the true minimum-phase filter and the one obtained using the DARE method, and from this it is possible to see that convergence of the two iterative methods is almost identical. In Table I the complexity of computing the minimum-phase filter using the two iterative methods has been compared (based on (7) and (9)), and from this it is seen that the fast QL-factorization method has a computational advantage.

⁷Again, strictly speaking the convergence occurs from the last row and up, since it is the QL-factorization.

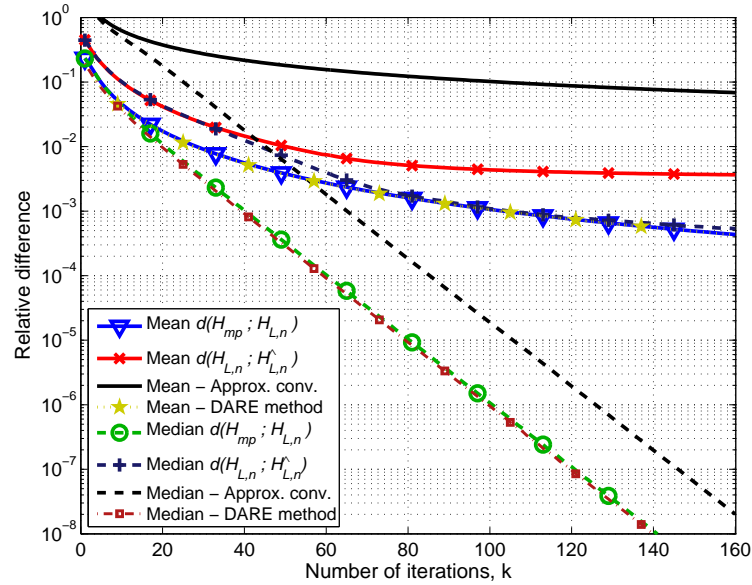


Figure 4. The relative deviations $d(\mathcal{H}_{mp}; \mathcal{H}_{L,k})$ and $d(\mathcal{H}_{L,k}; \mathcal{H}_{\hat{L},k})$ in a SISO channel with Gaussian coefficients having uniform power in the delay domain, $L = 6$.

Table I

COMPLEXITY OF COMPUTING THE MINIMUM-PHASE FILTER USING THE FAST QL-FACTORIZATION (QL) AND THE DARE METHOD (DARE) USING k ITERATIONS IN A LENGTH L SISO SYSTEM.

k	Method	$L = 5$	$L = 10$	$L = 15$	$L = 20$
10	QL	$3.18 \cdot 10^2$	$5.98 \cdot 10^2$	$9.03 \cdot 10^2$	$1.23 \cdot 10^3$
	DARE	$3.70 \cdot 10^2$	$1.48 \cdot 10^3$	$3.34 \cdot 10^3$	$5.95 \cdot 10^3$
20	QL	$6.38 \cdot 10^2$	$1.17 \cdot 10^3$	$1.72 \cdot 10^3$	$2.30 \cdot 10^3$
	DARE	$7.30 \cdot 10^2$	$2.94 \cdot 10^3$	$6.65 \cdot 10^3$	$1.19 \cdot 10^4$

In Fig. 5 the result for the TU0 profile is shown and it is seen that the convergence rate is faster for this channel type compared with the Gaussian filter coefficients with uniform power in the delay domain. It is again observed that the median value of the difference decreases more rapidly than the mean value. We also see that the approximated convergence of $d(\mathcal{H}_{mp}; \mathcal{H}_{L,k})$ is even closer to the actual convergence for this channel profile and that the DARE method again has similar convergence.

In Fig. 6 we have a plot of the location of the roots of a 2×2 MIMO system having Gaussian coefficients with filter length $L = 5$, leading to 8 roots. From the plot it is seen that the roots of $\mathbf{H}(z)$ (illustrated with squares) which lie outside the unit circle are reflected inside (the circles) using the root method. Furthermore, it is seen that these roots match the roots of $\mathbf{L}(z)$. In Fig. 7 the root difference $\Delta z_k^{(l)} =$

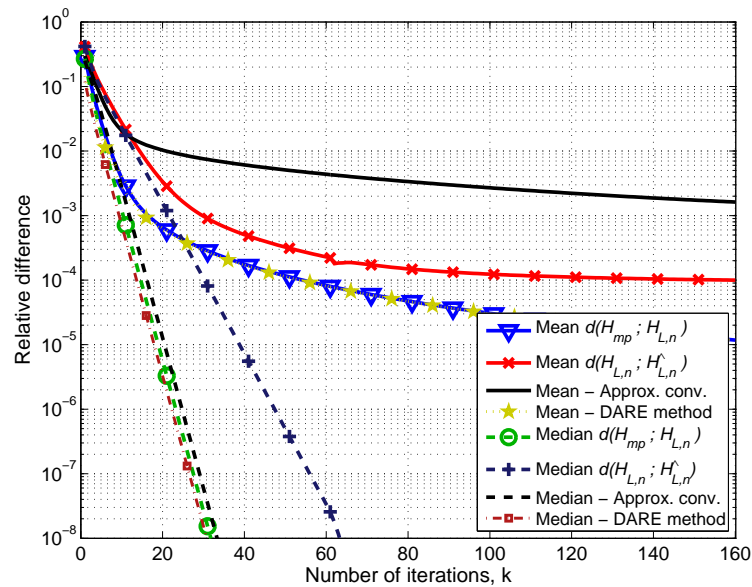


Figure 5. The relative deviations $d(\mathcal{H}_{mp}; \mathcal{H}_{L,k})$ and $d(\mathcal{H}_{L,k}; \mathcal{H}_{\hat{L},k})$ in the SISO channel TU0 with $L = 5$.

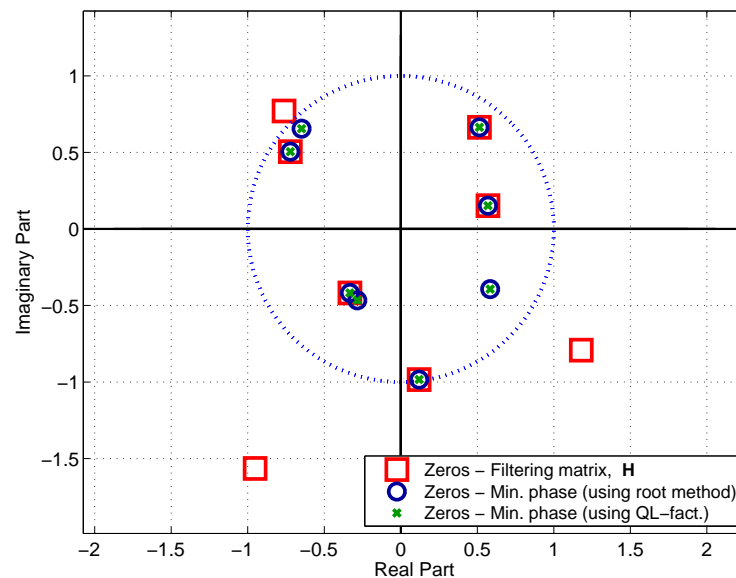


Figure 6. Locations of roots in a 2×2 MIMO system with Gaussian filter coefficients, $L = 5$. The number of iterations in the QL-factorization is $k = 200$.

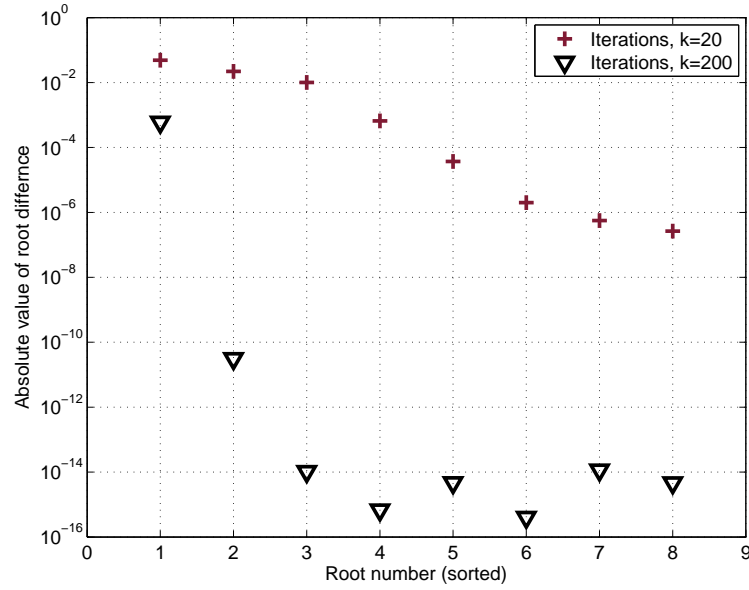


Figure 7. Deviation of the roots in a 2×2 MIMO system with Gaussian filter coefficients, $L = 5$ for iteration $k = 20$ and $k = 200$.

$z_{mp}^{(l)} - z_{L,k}^{(l)}$ has been plotted for each of the roots $l = \{1, \dots, 8\}$ for iteration $k = 20$ and $k = 200$. The roots have been sorted according to their distance to the unit circle, such that the one closest to the unit circle is called root 1, etc. The figure shows that the closer the root is to the unit circle, the slower the convergence it will have, which follows the convergence analysis given in Section V. After $k = 200$ iterations, it is primarily the root closest to the unit circle which contributes to the difference between the filter obtained from \mathbf{L} and the true minimum-phase filter.

VII. CONCLUSION

It has been shown how the QL-factorization of the filtering matrix gives the finite length equivalent to the minimum-phase and the all-pass filters and thereby presents a new way of computing these two classical filters in a numerically stable way. The exact convergence rate has been computed for a simple SISO length $L = 2$ system and an upper bound has been derived, which is used for approximating the convergence in systems of arbitrary length. Asymptotically these results also generalize to MIMO systems. The derived convergence results correspond well with what is observed in simulations and, furthermore, they are in agreement with existing results for the DARE method.

ACKNOWLEDGMENT

The authors would like to thank Prof. Babak Hassibi, for the useful discussions on spectral factorization and the Discrete-time Algebraic Riccati Equation. We also thank the reviewers for useful comments and suggestions.

APPENDIX A

FUNCTIONAL FORM OF RECURSIVE PARAMETER γ_k

In this appendix we rewrite the recursive expression for the parameter γ_k to a functional form. The recursion given in (15) is

$$\gamma_{k+1} = \frac{a |\gamma_k|}{\sqrt{1 + |\gamma_k|^2}} ,$$

and by inserting $\gamma_1 = a$ into the expression above we get $\gamma_2 = e^{i\angle a} |a|^2 / \sqrt{1 + |a|^2}$ and we then see that

$$\gamma_3 = \frac{e^{i\angle a} |a| \frac{|a|^2}{\sqrt{1+|a|^2}}}{\sqrt{1 + \frac{|a|^4}{1+|a|^2}}} = \frac{e^{i\angle a} |a|^3}{\sqrt{1 + |a|^2 + |a|^4}} . \quad (30)$$

The reduction made in (30) can be repeated in every step of the recursion and, therefore, in general we get

$$\gamma_k = \frac{e^{i\angle a} |a|^k}{\sqrt{\sum_{k'=0}^{k-1} |a|^{2k'}}} = \frac{e^{i\angle a} |a|^k}{\sqrt{\frac{1-|a|^{2(k-1)+2}}{1-|a|^2}}} \quad (31a)$$

$$= e^{i\angle a} |a|^k \sqrt{\frac{|a|^2 - 1}{|a|^{2k} - 1}} . \quad (31b)$$

APPENDIX B

FIXED-POINT SOLUTIONS

This appendix proves Lemma V.2. From Section V-A we have seen that the parameter γ_k , which determines the coefficients in \mathbf{L} of the Householder transformation, can be computed recursively. In order to find the fixed-point solution

$$\gamma_{fix} = f(\gamma_{fix}) , \quad \text{where } f(x) = \frac{a |x|}{\sqrt{1 + |x|^2}} , \quad (32)$$

we will first assume that $\gamma_{fix} \neq 0$. We then have

$$\gamma_{fix} = \frac{a |\gamma_{fix}|}{\sqrt{1 + |\gamma_{fix}|^2}} \Leftrightarrow \quad (33a)$$

$$a = \sqrt{1 + |\gamma_{fix}|^2} e^{i\angle\gamma_{fix}} \Leftrightarrow \quad (33b)$$

$$|a|^2 = 1 + |\gamma_{fix}|^2 \Rightarrow \quad (33c)$$

$$\gamma_{fix} = e^{i\angle a} \sqrt{|a|^2 - 1}, \quad \text{where } |a| > 1. \quad (33d)$$

In (33b) and (33d) we have used the fact that $\angle\gamma_{fix}$ will always be the same as $\angle a$. By inserting (33d) in (32) it is easily verified that this is indeed a fixed-point solution for $|a| > 1$.

Let us next consider the case where $\gamma_{fix} = 0$. By inserting this value of γ_{fix} into (32), it is easily seen that this is actually a fixed-point. Thus, Lemma V.2 has hereby been proved.

APPENDIX C

CONVERGENCE TO FIXED-POINT SOLUTIONS

In order to prove that the recursive update of γ_k converges to the fixed-points we are interested in showing that

$$|\gamma_{k+1} - \gamma_{fix}| < |\gamma_k - \gamma_{fix}|, \quad \forall k \geq 1. \quad (34)$$

The proof has been split up into the case for $|a| \leq 1$ and $|a| > 1$, but we will first prove Lemma C.1, which turns out to be a useful lemma when proving the convergence to the fixed-point for $|a| > 1$.

Lemma C.1. *Let γ_{fix} be the fixed-point solution given in Lemma V.2 and let ϵ be a complex valued constant with $\angle(\epsilon) = \angle\gamma_{fix} = \angle a$, we then have*

$$|f(\epsilon + \gamma_{fix})| > |\gamma_{fix}|, \quad \text{where } f(x) = \frac{a|x|}{\sqrt{1 + |x|^2}}. \quad (35)$$

Proof: It is trivial to show that Lemma C.1 is valid for $|a| \leq 1$, since $\gamma_{fix} = 0$, so we will focus on $\gamma_{fix} = e^{i\angle a} \sqrt{|a|^2 - 1}$. Since the arguments for all the terms in (35) are the same, we can simply ignore the angle and only consider the simple case where all values are real and positive, i.e.

$$f(\epsilon + \gamma_{fix}) > \gamma_{fix} \Rightarrow \quad (36a)$$

$$\frac{a(\epsilon + \sqrt{a^2 - 1})}{\sqrt{1 + (\epsilon + \sqrt{a^2 - 1})^2}} > \sqrt{a^2 - 1}, \quad (36b)$$

which can be rewritten as

$$\frac{1}{(\epsilon^2 + 2\epsilon\sqrt{a^2 - 1})/a^2 + 1} < 1. \quad (37)$$

Since all the terms on the LHS in (37) are positive, the inequality is true for all $|a| > 1$. This completes the proof. \blacksquare

We now turn our attention to proving (34).

A. Case where $|a| \leq 1$

It is fairly straight forward to prove the convergence in this case, since $\gamma_{fix} = 0$ and, thus, (34) reduces to showing that $\lim_{k \rightarrow \infty} \gamma_k = 0$, or equivalently $|\gamma_k| > |\gamma_{k+1}|$,

$$|\gamma_k| > |\gamma_{k+1}| = \frac{|a| |\gamma_k|}{\sqrt{1 + |\gamma_k|^2}} \Leftrightarrow \quad (38a)$$

$$\sqrt{1 + |\gamma_k|^2} > |a| . \quad (38b)$$

Since we assumed that $|a| \leq 1$, (38b) is satisfied for all $\gamma_{fix} \neq 0$. In the case where $\gamma = 0$ we have the fixed-point, and therefore, we have proved the convergence to the fixed-point for $|a| \leq 1$.

B. Case where $|a| > 1$

Since the initial input to the recursion in (15) is a , which numerically is greater than γ_{fix} , we can use Lemma C.1 to rewrite (34) as

$$|\gamma_{k+1}| - |\gamma_{fix}| < |\gamma_k| - |\gamma_{fix}| \Leftrightarrow \quad (39a)$$

$$|\gamma_{k+1}| = \frac{|a| |\gamma_k|}{\sqrt{1 + |\gamma_k|^2}} < |\gamma_k| \Leftrightarrow \quad (39b)$$

$$\frac{|a|}{\sqrt{1 + |\gamma_k|^2}} < 1 . \quad (39c)$$

By recalling from Lemma C.1 that $|\gamma_{fix}| < |\gamma_k|$ (when $\gamma_k \neq \gamma_{fix}$), we can upper bound LHS in (39c) as

$$\frac{|a|}{\sqrt{1 + |\gamma_k|^2}} < \frac{|a|}{\sqrt{1 + |\gamma_{fix}|^2}} \quad (40a)$$

$$= \frac{|a|}{\sqrt{1 + |a|^2 - 1}} = 1 . \quad (40b)$$

Now since the RHS in (40a) is equal to one, (39) must be true, which completes the proof of convergence for $|a| > 1$, and provides us with the following lemma.

Lemma C.2 (Convergence to fixed-point). *Let γ_{fix} be the fixed-point solution given in Lemma V.2 and let $\gamma_k \neq 0$ be a complex valued number given by the recursion in (15). For $\gamma_1 = a$ the value of γ_k satisfy*

$$|\gamma_{k+1} - \gamma_{fix}| < |\gamma_k - \gamma_{fix}| , \quad \forall k \geq 1 .$$

REFERENCES

- [1] J. G. Proakis and D. G. Manolakis, *Digital Signal Processing*, 3rd ed. Prentice Hall, 1995.
- [2] A. Oppenheim and R. Schaffer, *Discrete-Time Signal Processing*. Prentice-Hall, 1989.
- [3] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*. Prentice Hall, 2000.
- [4] A. Duel-Hallen, , and C. Heegard, "Delayed Decision-Feedback Sequence Estimation," *IEEE Trans. on Commun.*, vol. 37, pp. 428–436, May 1989.
- [5] M. V. Eyuboglu and S. U. H. Qureshi, "Reduced-State Sequence Estimation with Set Partitioning and Decision Feedback," *IEEE Trans. on Commun.*, vol. 36, pp. 13–20, Jan. 1988.
- [6] W. H. Gerstacker, F. Obernosterer, M. R., and J. B. Huber, "On Prefilter Computation for Reduced-State Equalization," *IEEE Trans. on Wireless Commun.*, vol. 1, pp. 793–800, Oct. 2002.
- [7] G. Golub and C. Van Loan, *Matrix Computations*. Johns Hopkins University Press, 1996.
- [8] R. Horn and C. Johnson, *Matrix Analysis*. Cambridge University Press, 1990.
- [9] M. Hansen and L. P. B. Christensen, "Efficient Minimum-Phase Prefilter Computation Using Fast QL-Factorization," in *ICASSP'09*, 2009.
- [10] A. W. Bojanczyk, R. P. Brent, and F. d. Hoog, "QR Factorization of Toeplitz Matrices," *Numer. Math.*, vol. 49, pp. 81–94, Jul. 1986.
- [11] D. R. Sweet, "Fast Toeplitz Orthogonalization," *Numer. Math.*, vol. 43, pp. 1–21, Feb. 1984.
- [12] S. Qiao, "Hybrid Algorithm for Fast Toeplitz Orthogonalization," *Numer. Math.*, vol. 53, pp. 351–366, May 1988.
- [13] D. R. Sweet, "Fast Block Toeplitz Orthogonalization," *Numer. Math.*, vol. 58, pp. 613–629, 1991.
- [14] N. Al-Dhahir and J. Cioffi, "Finite-length vs. infinite-length MMSE-DFE: The Connection," in *Signals, Systems and Computers, 1993. 1993 Conference Record of The Twenty-Seventh Asilomar Conference on*, 1993, pp. 677–681.
- [15] A. H. Sayed and T. Kailath, "A Survey of Spectral Factorization Methods," *Numerical Linear Algebra with Applications*, vol. 8, pp. 467–496, Jul. 2001.
- [16] V. Kucera, "Factorization of Rational Spectral Matrices: A Survey of Methods," in *International Conference on Control 1991.*, 1991, pp. 1074–1078.
- [17] J. F. Claerbout, *Fundamentals of Geophysical Data Processing*. Blackwell Scientific Publications, 1985.
- [18] D. Youla, "On the Factorization of Rational Matrices," *Information Theory, IRE Transactions on*, vol. 7, no. 3, pp. 172–189, 1961.
- [19] D. Youla and N. Kazanjian, "Bauer-Type Factorization of Positive Matrices and the Theory of Matrix Polynomials Orthogonal on the Unit Circle," *IEEE Trans. on Circuits and Systems*, vol. 25, no. 2, pp. 57–69, 1978.
- [20] Y. Rozanov, "Spectral Properties of Multivariate Stationary Processes and Boundary Properties of Analytic Matrices," *Theory of Probability and its Applications*, vol. 5, p. 362, 1960.
- [21] A. Yaglom, "Effective Solutions of Linear Approximation Problems for Multivariate Stationary Processes with a Rational Spectrum," *Theory of Probability and its Applications*, vol. 5, pp. 239–264, 1960.
- [22] R. F. H. Fischer, "Sorted Spectral Factorization of Matrix Polynomials in MIMO Communications," *IEEE Trans. on Info. Theory*, vol. 53, pp. 945–951, Jun. 2005.
- [23] J. E. Dennis Jr., J. F. Traub, and R. P. Weber, "The Algebraic Theory of Matrix Polynomials," *SIAM Journal on Numerical Analysis*, vol. 13, pp. 813–845, Dec. 1976.
- [24] —, "Algorithms for Solvents of Matrix Polynomials," *SIAM Journal on Numerical Analysis*, pp. 523–533, 1978.

- [25] B. Anderson, K. Hitz, and N. Diem, "Recursive Algorithm for Spectral Factorization," *IEEE Transactions on Circuits and Systems*, vol. 21, no. 6, pp. 742–750, 1974.
- [26] L. P. B. Christensen, "Signal Processing for Improved Wireless Receiver Performance," Ph.D. dissertation, Informatics and Mathematical Modelling, Technical University of Denmark, DTU, 2007.
- [27] M. Hansen, L. P. B. Christensen, and O. Winther, "On Sphere Detection and Minimum-Phase Prefiltered Reduced-State Sequence Estimation," in *GLOBECOM'07*, 2007, pp. 4237–4241.
- [28] N. Al-Dhahir and J. M. Cioffi, "MMSE Decision-Feedback Equalizers: Finite-Length Results," *IEEE Trans. on Info. Theory*, vol. 41, pp. 961–975, Jul. 1995.
- [29] R. Gray, *Toeplitz And Circulant Matrices: A Review (Foundations and Trends in Communications and Information Theory)*. Now Publishers Inc. Hanover, MA, USA, 2006.
- [30] T. I. Laakso and V. Välimäki, "Energy-Based Effective Length of the Impulse Response of a Recursive Filter," in *ICASSP'98*, vol. 3, 1998, pp. pp. 1253–1256.
- [31] 3GPP TS 45.005, *3GPP TSG GERAN. Radio Transmission and Reception* (Release 5).
- [32] P. Laurent, "Exact and Approximate Construction of Digital Phase Modulations by Superposition of Amplitude Modulated Pulses (AMP)," *IEEE Trans. on Commun.*, vol. 34, no. 2, pp. 150–160, 1986.
- [33] 3GPP TS 45.004, *3GPP TS GERAN. Radio Access Network; Modulation* (Release 8), 2008.



Morten Hansen (S'07, M'09) was born in Copenhagen, Denmark, in 1979. He received his B.E. and M.Sc. degrees in electrical engineering from the Technical University of Denmark (DTU), in 2004 and 2006, respectively. He is currently pursuing his Ph.D. degree in signal processing for wireless communications at DTU. During his Ph.D. studies he has been a graduate student visitor at California Institute of Technology (CalTech) in Prof. Babak Hassibi's group.



Lars P. B. Christensen (S'05, M'08) received his M.Sc. and Ph.D. degrees in 2003 and 2007 respectively, both in electrical engineering from the Technical University of Denmark. Since graduating he has been with the Modem Algorithm Design group at Nokia working on research and design of signal processing algorithms for wireless communications. His interests are within estimation, detection and decoding related to communication systems as well as the general areas of signal processing and information theory.



Ole Winther (OW) works in machine learning with applications in bioinformatics, data mining, 3G+ wireless communication and collaborative filtering. Ole Winther (M.Sc. physics and computer science '94 and Ph.D. physics '98, both at University of Copenhagen (KU)) has published 50+ scientific papers and has previously held positions at Lund University and Center for Biological Sequence Analysis, Technical University of Denmark (DTU). He currently holds joint positions as associate professor at Intelligent Signal Processing (ISP), IMM, DTU and group leader at Bioinformatics, KU.